

VU Research Portal

Harmonizing the Dutch Historical Censuses in the Semantic Web

Ashkpour, A.; Merono, A.

2014

document version

Early version, also known as pre-print

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Ashkpour, A., & Merono, A. (2014). *Harmonizing the Dutch Historical Censuses in the Semantic Web*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

The CEDAR Project: Harmonizing the Dutch Historical censuses in the Semantic Web

Ashkan Ashkpour – International Institute of Social History, E-humanities Group, Erasmus university

Albert Meroño-Peñuela – Data Archiving and Networked Services, E-Humanities Group, VU University Amsterdam

Historical censuses are on the most consulted, reliable and large scale statistical data sources available which give an insight in the population characteristics of a nation: they provide a wealth of data on many issues in the course of time at the demographic, social and economic level. In the Netherlands, the digitization of the Dutch historical censuses (1795-1971) has resulted in many dispersed tables, and meaningful historical information is currently hidden in aggregated data over 2,300 of these tables. In order to fully reap the benefits of this dataset temporal comparisons are required. However, the Dutch historical censuses are still very difficult to compare, aggregate and query in a uniform fashion due to the lack of data harmonization. The CEDAR project of the Computational Humanities Programme aims at enabling greater access and use of this dataset by applying a specific datamodel (exploiting the Resource Description Framework RDF technology), to create Linked Census Data and make it interlinkable with other hubs of historical socioeconomic and demographic information; and various harmonization practices. The process of harmonization is currently an ongoing process and will be so in the future. In order to implement harmonization on this dataset we have applied a three tier model, distinguishing between the raw data layer (containing a direct translation of the numbers and concepts in the tables), the annotations layer (with corrections and comments from the dataset curators), and the harmonization layer (linking and standardizing heterogeneous census entities). Making extensive use of semantic technologies and Linked Data principles, we make this Linked Census Data comparable over time by leveraging specific data classifications, taxonomies and ontologies. By querying these, we create visualizations in order to explore the thousands of variables and to get an early understanding of our data and its relations. We create bottom up classifications for information contained in all three census types (population, houses and occupations) like demographical structures, housing types, occupational classes and statuses, religious denominations, and so on. We use animation techniques to display the conceptual changes that modified the social landscape in fundamental centuries of Europe's history.